

Henry Ford Health System

## Henry Ford Health System Scholarly Commons

---

Center for Health Policy and Health Services  
Research Articles

Center for Health Policy and Health Services  
Research

---

1-1-2017

### Validity of an automated algorithm using diagnosis and procedure codes to identify decompensated cirrhosis using electronic health records

Mei Lu

*Henry Ford Health System, mlu1@hfhs.org*

Wadih Chacra

David Rabin

Loralee B. Rupp

*Henry Ford Health System, lrupp1@hfhs.org*

Sheri Trudeau

*Henry Ford Health System, STRUDEA1@hfhs.org*

*See next page for additional authors*

Follow this and additional works at: [https://scholarlycommons.henryford.com/chphsr\\_articles](https://scholarlycommons.henryford.com/chphsr_articles)

---

#### Recommended Citation

Lu M, Chacra W, Rabin D, Rupp LB, Trudeau S, Li J, Gordon SC. Validity of an automated algorithm using diagnosis and procedure codes to identify decompensated cirrhosis using electronic health records. Clin Epidemiol 2017; 9:369-376.

This Article is brought to you for free and open access by the Center for Health Policy and Health Services Research at Henry Ford Health System Scholarly Commons. It has been accepted for inclusion in Center for Health Policy and Health Services Research Articles by an authorized administrator of Henry Ford Health System Scholarly Commons.

---

## Authors

Mei Lu, Wadih Chacra, David Rabin, Lorelee B. Rupp, Sheri Trudeau, Jia Li, and Stuart C. Gordon

# Validity of an automated algorithm using diagnosis and procedure codes to identify decompensated cirrhosis using electronic health records

Mei Lu<sup>1</sup>  
Wadih Chacra<sup>2</sup>  
David Rabin<sup>3</sup>  
Loralee B Rupp<sup>4</sup>  
Sheri Trudeau<sup>1</sup>  
Jia Li<sup>1</sup>  
Stuart C Gordon<sup>5</sup>

## On behalf of the Chronic Hepatitis Cohort Study (CHeCS) Investigators

<sup>1</sup>Department of Public Health Sciences, Henry Ford Health System, Detroit, MI, USA; <sup>2</sup>Division of Gastroenterology and Hepatology, University of Illinois College of Medicine, Chicago, IL, USA; <sup>3</sup>Atlanta Gastroenterology Associates, Atlanta, GA, USA; <sup>4</sup>Center for Health Policy & Health Services Research, Henry Ford Health System, Detroit MI, USA; <sup>5</sup>Division of Gastroenterology and Hepatology, Henry Ford Health System, Detroit, MI, USA

Correspondence: Mei Lu  
Department of Public Health Sciences,  
Henry Ford Health System, 3E One Ford  
Place, Detroit, MI, USA  
Tel +1 313 874 6413  
Fax +1 313 874 6730  
Email mlu1@hfhs.org

**Abstract:** Viral hepatitis-induced cirrhosis can progress to decompensated cirrhosis. Clinical decompensation represents a milestone event for the patient with cirrhosis, yet there remains uncertainty regarding precisely how to define this important phenomenon. With the development of broader treatment options for cirrhotic hepatitis patients, efficient identification of liver status before evolving to decompensated cirrhosis could be life-saving, but research on the topic has been limited by inconsistencies across studies, populations, and case-confirmation methods. We sought to determine whether diagnosis/procedure codes drawn from electronic health records (EHRs) could be used to identify patients with decompensated cirrhosis. In our first step, chart review was used to determine liver status (compensated cirrhosis, decompensated cirrhosis, non-cirrhotic) in patients from the Chronic Hepatitis Cohort Study. Next, a hybrid approach between Least Absolute Shrinkage and Selection Operator regression and Classification Regression Trees models was used to optimize EHR-based identification of decompensated cirrhosis, based on 41 diagnosis and procedure codes. These models were validated using tenfold cross-validation; method accuracy was evaluated by positive predictive values (PPVs) and area under receiver operating characteristic (AUROC) curves. Among 296 patients (23 with hepatitis B, 268 with hepatitis C, and 5 co-infected) with a 2:1 ratio of biopsy-confirmed cirrhosis to noncirrhosis, chart review identified 127 cases of decompensated cirrhosis (Kappa=0.88). The algorithm of five liver-related conditions—liver transplant, hepatocellular carcinoma, esophageal varices complications/procedures, ascites, and cirrhosis—yielded a PPV of 85% and an AUROC of 92%. A hierarchical subset of three conditions (hepatocellular carcinoma, ascites, and esophageal varices) demonstrated a PPV of 81% and an AUROC of 86%. Given the excellent predictive ability of our model, this EHR-based automated algorithm may be used to successfully identify patients with decompensated cirrhosis. This algorithm may contribute to timely identification and treatment of viral hepatitis patients who have progressed to decompensated cirrhosis.

**Keywords:** chronic viral hepatitis, hepatitis B, HBV, hepatitis C, HCV, classification and regression tree modeling, CART modeling

## Introduction

Complications of liver cirrhosis include an increased risk of hepatocellular carcinoma (HCC), liver transplant, and death from liver failure.<sup>1–3</sup> In “compensated” cirrhosis, the damaged liver still functions adequately; cirrhosis that has progressed sufficiently to interfere with essential bodily functions is classified as “decompensated”. The development of clinical decompensation represents a milestone event for the patient with cirrhosis, and is associated with increased morbidity, mortality, and utilization of

health care resources—yet there remains uncertainty regarding precisely how to define this important phenomenon.

Although liver biopsy is considered the gold standard for confirming the presence of cirrhosis, it cannot be used to determine hepatic functional status.<sup>4,5</sup> Measurement of hepatic venous pressure gradient has been shown to be an excellent predictor of stage of fibrosis and risk for decompensated cirrhosis (DC), but the procedure carries risks and is not commonly performed in clinical practice. While the Child-Pugh classification stages severity of cirrhosis and the Model for End-Stage Liver Disease (MELD) scores predict survival in a patient with cirrhosis, there remains considerable ambiguity regarding whether a cirrhotic patient has experienced decompensation.<sup>6</sup> Signs and symptoms (including past or present ascites, encephalopathy, and even gastrointestinal bleeding) represent continuums that may be variously interpreted. In clinical trials, for example, decompensated cirrhotics are categorized as distinct cohorts. An efficient system to identify decompensated patients from administrative data may inform early-intervention clinical trials or epidemiological research to reduce poor outcomes.

Epidemiological, outcome, and economic studies of viral hepatitis often rely on International Classification of Diseases, 9th (or 10th) Revision (ICD-9/10) and Current Procedural Terminology (CPT) codes obtained from administrative, claims, or electronic health record (EHR) databases to identify patients with cirrhosis and decompensation.<sup>3,7,8</sup> However, diagnostic and procedure code data are subject to limitations, including coding inaccuracy and incompleteness. Several studies have been performed in United States Veterans Administration populations<sup>1,9,10</sup> to evaluate the validity of diagnosis codes for identification of cirrhosis/DC (alone or in combination with laboratory markers); however, these studies were limited by use of a restricted patient population, inconsistent results, or a focus on risk factors associated with DC rather than disease classification. A more recent study in a civilian population validated a model for identifying cirrhosis (a classification study) based on diagnosis codes and laboratory abnormalities.<sup>11</sup> However, the outcome in this study was “cirrhosis”, not specifically decompensation; likewise, laboratory values are not typically available in administrative and claims databases (for example, EHR data). Therefore, we sought to develop and optimize a model based exclusively on electronic data-derived ICD-9 and CPT codes to accurately identify patients with DC in a civilian cohort of viral hepatitis patients.

## Materials and methods

### Study design

Our sample was drawn from a subset of hepatitis patients from the Chronic Hepatitis Cohort Study (CHeCS), an observational study of patients from four large US health care systems, including Henry Ford Health System (HFHS, Detroit, MI). The study follows all guidelines of the US Department of Health and Human Services regarding the protection of human subjects; our protocol was approved and renewed annually by the Institutional Review Boards of HFHS; Geisinger Health System; Kaiser Permanente Northwest; and Kaiser Permanente Hawai'i. Due to the de-identified nature of this observational study, requirements for written informed consent were waived. Patient data were anonymized and de-identified prior to analysis. HFHS CHeCS patients were considered for study inclusion if they had biopsy-confirmed or excluded cirrhosis and clinical encounters data collected through the earlier date of liver transplant (if applicable) or December 31, 2010. In patients with multiple biopsies, the most severe biopsy stage was used. A random sample of 296 patients with a 2:1 ratio of cirrhosis and noncirrhosis was selected using SAS.<sup>12</sup> Harrell's 1-to-10 rule of thumb<sup>13</sup> was used to estimate sample size; assuming that 50% of cirrhosis cases were decompensated, up to 10 binary/continuous predictors for multivariable modeling could be analyzed without overfitting the data.

Chart reviews were performed to classify the patients into one of three categories: 1) noncirrhotic (NC); 2) compensated cirrhotic (CC); and 3) decompensated cirrhotic (DC). To ensure consistency, a senior hepatologist (SG) provided standard diagnosis guidelines and two gastroenterology fellows (“raters”) performed the chart reviews. Confirmation of at least one of the following conditions during chart review resulted in a diagnosis of DC: hepatic encephalopathy; gastrointestinal bleeding due to portal hypertension; jaundice with a total bilirubin of >2.5 mg/dL; ascites/hydrothorax due to portal hypertension; or HCC. The two raters independently reviewed all patient charts. Weighted kappa ( $\kappa$ ) was used to measure the agreement between the raters, accounting for the ordinal property of cirrhosis levels, where  $\kappa > 0.80$  indicates “excellent” agreement. Any discrepancies in classification between the raters were reviewed by the senior hepatologist for adjudication. Final clinical classification of DC versus compensated cirrhosis (CC) or noncirrhosis was used as the outcome of interest.

We hypothesized that a set of diagnosis and procedure codes could identify DC. To begin, we modified a set of 41

ICD-9 and CPT codes previously used by Spradling et al<sup>7</sup> to identify cirrhotic status. These codes were grouped into ten clusters related to clinical presentation of hepatic decompensation (Table 1). The data collected from each cluster were either quantitative (counts) or categorical (presence/absence). In the previous study,<sup>7</sup> a three-level cirrhosis classification variable (NC, CC, and DC) was derived from ten disease conditions (clusters). The variable was defined as “NC” if no ICD-9/CPT codes occurred in any of the C1–10 cluster categories, as “CC” if any codes from the C10 cluster were present and codes from the C1–C9 clusters were absent, and as “DC” if any codes from the C1–C9 clusters were present. We first validated this three-level cirrhosis classification variable, then developed and optimized models to predict DC based on the individual C1–C10 cluster variables for greater accuracy.

## Statistical analysis

Descriptive statistics were used to summarize characteristics of the study population. Mean and standard deviation were used for continuous variables; frequency and percentage were used for categorical variables. Two novel classifiers were considered for the classification of DC; the first was logistic regression using the Least Absolute Shrinkage and Selection Operator (LASSO)<sup>14</sup> approach for variable selection

that yields a variable combination with optimal model predictive ability. The second model used Classification and Regression Trees (CART),<sup>15,16</sup> a nonparametric approach that generates a binary recursive multilevel tree that partitions the variables most predictive of the outcome of interest (in this case, DC).<sup>14</sup> The CART method is ideally suited for clinical decision making because it can reveal important relationships between variables that may remain hidden when using logistic regression.<sup>15,16</sup> Finally, a hybrid method between generalized LASSO and CART was used to simplify the model for clinical interpretation without loss of generalizability and accuracy.

Model learning and testing were performed based on the sample of 296 patients. Tenfold cross-validation was performed and model predictive accuracy was based on validation results. Predictive accuracy of the model was assessed using area under receiver operating characteristic (AUROC) curves, as well as positive predictive value (PPV), sensitivity, and specificity. Models are considered to have “reasonable” to “excellent” predictive ability when the AUROC is 70%–80% and 80%–90%, respectively. The nonparametric Mann–Whitney approach<sup>17</sup> was used to compare the difference in AUROC between models. All analyses were performed using SPM version 8.0 (Salford Predictive Modeler)<sup>14</sup> and SAS version 9.4.<sup>12</sup>

**Table 1** ICD-9/10 and CPT codes comprising the 10 cluster variables (C1–C10)

Condition	Associated ICD-9 and CPT codes	Associated ICD-9/10 and CPT codes
C1: Liver transplant	V42.7, 996.82, 050.5, 050.51, 050.59, 47135, 47136	V42.7, 996.82, 050.5, 050.51, 050.59, 47135, 47136, T86.40, T86.41, T86.42, 0FY00Z0, 0FY00Z1, 0FY00Z2
C2: Liver cancer	155.0, 155.1, 155.2	155.0, 155.1, 155.2, C22.0, C22.2, C22.7, C22.8, C22.1, C22.9
C3: Hepatorenal syndrome	572.4	572.4, K76.7
C4: Hepatic encephalopathy	572.2	572.2, K72.09, K72.91
C5: Portal hypertension/portal decompression procedures	572.3, 37140, 37160, 37180, 37181, 37182, 37183	572.3, 37140, 37160, 37180, 37181, 37182, 37183, K76.6
C6: Esophageal varices complications (bleeding) and procedures	456.0, 456.20, 43204, 43205, 43243, 43244, 43400, 43401, 42.91, 44.91, 96.06	456.0, 456.20, 42.91, 44.91, 96.06, 43204, 43205, 43243, 43244, 43400, 43401, I85.01, I85.11, 06L30CZ, 06L30DZ, 06L30ZZ, 06L33CZ, 06L33DZ, 06L33ZZ, 06L34CZ, 06L34DZ, 06L34ZZ, 06L20ZZ, 06L23ZZ, 06L24ZZ, 0DL57DZ, 0DL58DZ
C7: Other gastrointestinal hemorrhage	530.7, 530.82, 578.0, 578.1, 578.9	530.7, 530.82, 578.0, 578.1, 578.9, K22.6, K22.8, K92.0, K92.1, K92.2
C8: Ascites/paracentesis procedures	789.5, 789.59, 49080, 49081, 54.91	789.5, 789.59, 54.91, 49080, 49081, R18.8, 0D9S30Z, 0D9S3ZZ, 0D9S40Z, 0D9S4ZZ, 0D9T30Z, 0D9T3ZZ, 0D9T40Z, 0D9T4ZZ, 0D9V30Z, 0D9V3ZZ, 0D9V40Z, 0D9V4ZZ, 0D9W30Z, 0D9W3ZZ, 0D9W40Z, 0D9W4ZZ, 0W9F30Z, 0W9F3ZZ, 0W9F40Z, 0W9F4ZZ, 0W9G30Z, 0W9G3ZZ, 0W9G40Z, 0W9G4ZZ, 0W9J30Z, 0W9J3ZZ
C9: Other sequelae of chronic liver disease	572.8	572.8, K72.10, K72.90
C10: Cirrhosis	571.2, 571.5	571.2, 571.5, K70.30, K74.0, K74.60, K74.69

**Abbreviations:** CPT, Current Procedural Terminology; ICD, International Classification of Diseases.

Given the ongoing implementation of ICD version 10, we conducted an ad hoc analysis based on both ICD-9 and ICD-10 liver-related diagnosis and procedure code data collected from 01/01/2015 to 10/01/2015, and compared C1–10 cluster variables as defined by both ICD-9 and ICD-9/10 codes. The detailed cluster variables with related ICD-10 codes are illustrated in Table 1.

## Results

Demographic characteristics are presented in Table 2. Two physician raters completed independent chart reviews to classify 296 patients (23 hepatitis B virus [HBV] infected, 268

hepatitis C virus [HCV] infected, and 5 co-infected). Agreement between our two raters was “excellent” ( $\kappa=0.88$ ; 95% Confidence Interval (CI): 0.84–0.92), indicating consistent and reliable classification assessment.<sup>18</sup> Eighteen inconsistent classifications were further reviewed by the senior hepatologist (SG). Based on this chart review, 102 (34%) patients did not have cirrhosis, 67 (23%) had CC, and 127 (42%) had DC. The previously defined three-level cirrhosis classification variable resulted in an AUROC of 80%, PPV of 70%, sensitivity of 90%, and specificity of 70% (data not shown).

## LASSO model

The inclusion of the full set (C1–C10) of ICD-9 variable clusters in the initial multivariate logistic LASSO regression model resulted in five retained clusters (C1, C2, C6, C8, and C10). Based on this model, the likelihood of DC can be predicted by a linear combination of the five cluster indicators:  $-1.52+0.98\times C1+1.17\times C2+0.98\times C6+1.25\times C8+0.48\times C10$ . Tenfold cross-validation estimated an AUROC of 89%, PPV of 83%, sensitivity of 83%, and specificity of 88%, using an optimized cutoff of 2.28 (Table 3A). The same predictive ability was observed when the cluster variables were quantitative (counts) or categorical (presence/absence).

For clinical applications, we then developed a single, simplified five-cluster indicator combination model,  $Z_1=C1+C2+C6+C8+C10$ , with  $Z_1$  ranging from 0 to 5, where higher scores indicate a greater likelihood of DC. The overall prediction accuracy (AUROC) remained similar, at 90% with tenfold validation (Table 3B). There was no difference between the AUROCs of the original and simplified models ( $p=0.09$ ).

## CART model

CART analysis generated a recursive tree with three cluster indicator variables (C2, C6, C8) and four terminal nodes (TN) (Figure 1). The AUROC was 86.8% using the learning data and 85.8% using the testing data (ten-fold validation). As illustrated in Figure 1, presence of HCC (C2) first identified 94 patients as decompensated (TN4, with 20 misclassified, 21%); patients without HCC would be identified as decompensated if there was indication of ascites (C8;  $n=35$  in TN3 with 5 misclassified, 14%); patients without either HCC or ascites would be identified as decompensated if they had bleeding esophageal varices (C6;  $n=6$  in TN2, with no misclassifications); PPV, sensitivity, and specificity were 81%, 87%, and 85%, respectively, for both the learning data and testing data (Table 3C). Notably, the three cluster variables

**Table 2** Study population characteristics

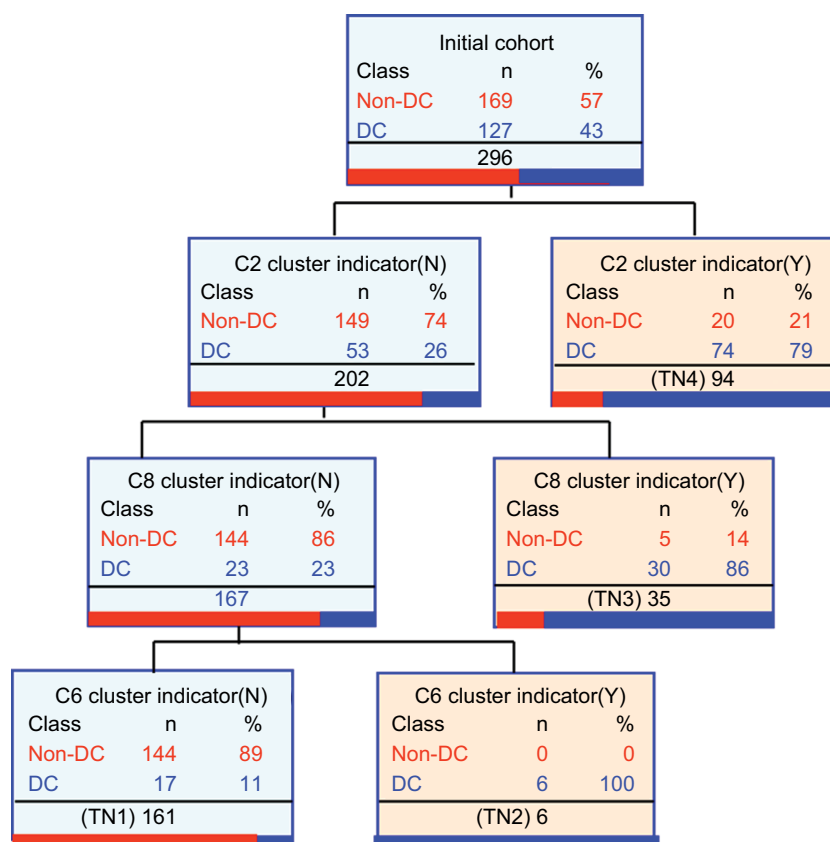
Characteristic	N=296
Age (mean, standard deviation)	58.6 (8.3)
Age category, (n)	
30≤40 years	4 (1%)
41≤50 years	27 (9%)
51≤60 years	161 (54%)
61≤70 years	75 (25%)
>70 years	29 (10%)
Male sex, (n)	175 (59%)
Race, (n)	
Asian/Pacific Islander	16 (5%)
White	162 (55%)
Black	108 (37%)
Native American	4 (1%)
Unknown	6 (2%)
Hispanic ethnicity, (n)	9 (3%)
Median annual household income, <sup>a</sup> (n)	
<\$15,000	6 (2%)
\$15,000<\$30,000	49 (17%)
\$30,000<\$50,000	106 (36%)
\$50,000<\$75,000	79 (27%)
≥\$75,000	47 (16%)
Missing	9 (3%)
Insurance status, (n)	
Medicaid	20 (7%)
Medicare	133 (45%)
Private	139 (47%)
None	4 (1%)
Unknown	334 (3%)
Liver conditions, (n)	
C1: Liver transplant	54 (18%)
C2: Liver cancer	94 (32%)
C3: Hepatorenal syndrome	11 (4%)
C4: Hepatic encephalopathy	30 (10%)
C5: Portal hypertension/ portal decompression procedures	25 (8%)
C6: Esophageal varices complications (bleeding) and procedures	24 (8%)
C7: Other gastrointestinal hemorrhage	53 (18%)
C8: Ascites/paracentesis procedures	65 (22%)
C9: Other sequelae of chronic liver disease	21 (7%)
C10: Cirrhosis	203 (67%)

**Notes:** <sup>a</sup>Currency is US\$.

**Table 3** Classification models for identification of DC

	(A) Optimized LASSO				(B) Five-cluster LASSO			(C) Three-cluster CART			(D) Five-cluster Hybrid		
Model	$-1.52+0.98\times C1+1.17\times C2+0.98\times C6+1.25\times C8+0.48\times C10$				$C1+C2+C6+C8+C10$ (with all coefficients=1)			C2, C6, C8 (Figure 1)			Combined LASSO/CART $\geq 2$ of C1, C2, C6, C8, and C10 (Figure 2)		
Cutoff	2.28				1.77						2		
Actual class	Non-DC	DC	Correct		Non-DC	DC	Correct	Non-DC	DC	Correct	Non-DC	DC	Correct
Testing data/model validation													
Non-DC	169	148	21	88%	145	24	86%	144	25	85%	145	24	86%
DC	127	21	106	83%	23	104	82%	17	110	87%	19	108	85%
Total	296	169	127		168	128		161	135		164	132	
AUROC		89%				90%			86%			86%	
Specificity		88%				86%			85%			86%	
Sensitivity		83%				82%			87%			85%	
PPV		83%				81%			81%			82%	
Learning Data/ Model Development													
Non-DC	169	151	18	89%	145	24	86%	143	26	85%	145	24	86%
DC	127	21	106	83%	19	108	85%	17	110	87%	19	108	85%
Total	296	172	124		164	132		160	136		164	132	
AUROC		92%				90%			87%			85%	
Specificity		89%				86%			85%			86%	
Sensitivity		83%				85%			87%			85%	
PPV		85%				82%			81%			82%	

**Abbreviations:** C1, liver transplant; C2, liver cancer; C6, esophageal varices complications (bleeding) and procedures; C8, ascites/paracentesis procedures; C10, cirrhosis; AUROC, area under receiver operating characteristic; CART, Classification and Regression Trees; DC, decompensated cirrhosis; LASSO, Least Absolute Shrinkage and Selection Operator; PPV, positive predictive value.

**Figure 1** The three cluster-variable CART model.

**Abbreviations:** C1, liver transplant; C2, liver cancer; C6, esophageal varices complications (bleeding) and procedures; C8, ascites/paracentesis procedures; C10, cirrhosis; CART, Classification and Regression Trees; DC, decompensated cirrhosis; TN, terminal node.



included in the final CART model were also identified in the LASSO regression model.

By using the CART approach with the simplified five-cluster combination variable ( $Z_1$ ) added to the ten cluster indicators (a hybrid method between LASSO and CART), only  $Z_1$  remained in the final CART model (Figure 2). A cutoff value of  $Z_1 \geq 2$  (indicating presence of codes from two or more of the five cluster variables) had a validated AUROC of 86%, PPV of 82%, sensitivity of 85%, and specificity of 86% (Table 3D).

In addition, a total of 2177 patients (1982 HCV, 116 HBV, and 18 HBV/HCV coinfecting) were considered who had at least one of the C1–C10-related diagnosis codes from 01/01/2015 to 10/01/2016. Of them, 125 patients had C1–C10-related ICD-10 diagnosis and procedures. There was complete concordance between ICD-9 and ICD-9/10 classifications for all patients except one, who presented with C2 (HCC)-related ICD-10 codes.

## Discussion

In clinical practice, the terms “decompensated cirrhosis” and “end-stage liver disease” lack precise definitions. Moreover, patients with cirrhosis may experience “episodes of decompensation”, such as transient ascites formation, and then revert back to a fully compensated state once ascites resolves. Similar events of variceal bleeding or hepatic encephalopathy, for example, may represent transient phenomena. With the advent of newer direct-acting all-oral (DAA) treatments for HCV, it is especially important to have consensus regarding this clinically important condition, because treatment options may vary if a cirrhotic patient has ever experienced “decompensation”. For example, the US Food and Drug Administration recently issued a warning regarding the use of protease

inhibitor antiviral therapy in chronic HCV patients who have ever experienced hepatic decompensation.

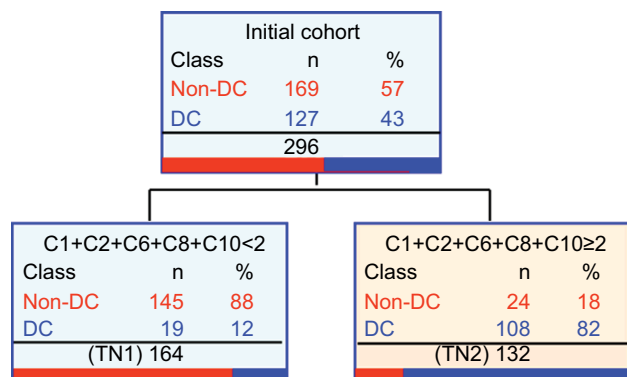
Our method—based on EHR-based diagnosis and procedure codes—may provide both clinicians and researchers an efficient and accurate method for the identification of decompensation among patients with viral hepatitis. Due to the low accuracy of individual diagnosis codes in detecting DC,<sup>7</sup> we grouped these codes into clusters of diagnoses related to specific manifestations of decompensation. We first validated a previously defined three-level classification variable based on ten clusters of 41 ICD-9 and CPT codes. We then developed an optimal algorithm based on the individual C1–C10 cluster variables for greater predictive ability.

To ensure the reliability of our classification of DC, medical records were reviewed by two trained raters and discrepancies were adjudicated by an expert hepatologist; agreement between the reviewers was “excellent” ( $\kappa=0.88$ ; 95% CI: 0.84–0.92), indicating high reliability. Our results showed that the predictive ability of the previously defined three-level cirrhosis classification variable (from 41 diagnosis and procedure codes) was “excellent” (AUROC=80%), although specificity was only “moderate” (70%), indicating a high false-positive rate.

The parametric LASSO regression model ( $-1.52+0.98 \times C1+1.17 \times C2+0.98 \times C6+1.25 \times C8+0.48 \times C10$ ; Table 3A), based on five cluster variables derived from 27 individual diagnosis and procedure codes, demonstrated the likelihood (probability) of decompensation on a continuous scale. The model’s predictive ability was in the upper range of “excellent” (AUROC=89%; PPV=83%). The simplified LASSO regression model (Table 3B) using a coefficient of 1 for all five cluster variables ( $Z_1=C1+C2+C6+C8+C10$ ) retained similar identification accuracy (AUROC=86%; PPV=82%). Notably, the nonparametric CART model ended with the same five cluster variables (Table 3C) obtained from the LASSO model, which indicates robustness of the DC identification process.

The classification model using a hybrid LASSO/CART approach with a cutoff of  $Z_1 \geq 2$  (Table 3D) retained similar identification accuracy (AUROC=86%; PPV=82%). Cirrhotic patients would be classified as “decompensated” if they ever had one of the following four conditions—liver transplant, HCC, complications or procedures related to esophageal varices, or ascites. This model may be more useful in a clinical setting or for researchers.

The abbreviated three-cluster variable model (derived from 17 individual diagnosis and procedure codes) identified DC (Figure 1) with only slightly reduced predictive accuracy (AUROC=86%; PPV=81%). Patients would be classified as “decompensated” at Level 1 in the presence of HCC-related



**Figure 2** The hybrid CART model.

**Abbreviations:** C1, liver transplant; C2, liver cancer; C6, esophageal varices, complications (bleeding), and procedures; C8, ascites/paracentesis procedures; CART, Classification and Regression Trees; DC, decompensated cirrhosis; TN, terminal node.



codes (C2); at Level 2 in the absence of HCC-related codes, but in the presence of ascites-related codes (C8); and at Level 3 in the absence of HCC- and ascites-related codes, but in the presence of codes related to bleeding esophageal varices (C6). The agreement between the hybrid five-cluster variable model with  $Z_1 \geq 2$  as the cutoff and the CART three-cluster models was “outstanding” ( $\kappa=0.93$ ; 95% CI: 0.88–0.97). This simplified model would be easy to use in clinical practice: cirrhotic patients may reliably be considered “ever decompensated” if they have one of the following four conditions in their medical records—liver transplant, HCC, complications or procedures related to esophageal varices, or ascites.

We note that our sample included only patients with chronic viral hepatitis. However, diagnosis and procedure codes related to viral hepatitis were not included in the model and, thus, were not relevant to its accuracy. Likewise, the conditions that define cirrhotic decompensation are the same regardless of etiology. This algorithm also requires further validation in an independent cohort with similar patient characteristics, as well as among patients with other causes of liver disease.

Our ad hoc analysis showed that ICD-10 codes enrich ICD-9 codes. In CART models using either the cutoff value of 1.77 or  $\geq 2$  C1–C10 clusters, we conclude that our ICD-9–based CART model could be adapted and replaced with ICD-9/10 codes.

In conclusion, our study showed that the presence of diagnosis and procedure codes in patient records could be used to efficiently identify patients with DC. We expect the results of this study to be useful for epidemiological and economic research of viral hepatitis-related liver disease, as well as serve as a quality improvement tool to ensure the best available care for such patients. Our algorithm may also be useful in the identification of risk factors for decompensation in observational studies with large cohorts of viral hepatitis patients. There are now effective therapies that may be life-saving for cirrhotic patients before or even after their condition worsens<sup>6</sup>—this tool may help clinicians recognize that cirrhosis has evolved to the point of decompensation. Our algorithm could be applied to electronic medical record systems to trigger automated alerts for specialty referrals; such alerts may contribute to timely clinical management of cirrhosis and its complications.

## Acknowledgments

The CHeCS Investigators include the following investigators and sites: Scott D Holmberg, Eyasu H Teshale, Philip

R Spradling, Anne C Moorman, Jim Xing, and Yuna Zhong, Division of Viral Hepatitis, National Centers for HIV, Viral Hepatitis, STD, and TB Prevention (NCHHSTP), Centers for Disease Control and Prevention (CDC), Atlanta, GA; Stuart C Gordon, David R Nerenz, Mei Lu, Lois Lamerato, Jia Li, Loralee B Rupp, Nonna Akkerman, Nancy Oja-Tebbe, Sheri Trudeau, Yueren Zhou, and Talan Zhang, Henry Ford Health System, Detroit, MI; Joseph A Boscarino, Zahra S Daar, and Robert E Smith, Center for Health Research, Geisinger Health System, Danville, PA; Yihe G Daida, Connie Mah Trinacty, Jonathan W Lai, Carmen P Wong, The Center for Health Research, Kaiser Permanente-Hawaii, Honolulu, HI; Mark A Schmidt and Judy L Donald, The Center for Health Research, Kaiser Permanente-Northwest, Portland, OR, USA.

Financial Support: Henry Ford Health System receives funding for CHeCS from the Centers for Disease Control and Prevention and from Gilead Sciences. CHeCS was previously funded through May 2016 by the CDC Foundation, which received grants from AbbVie; Genentech, A Member of the Roche Group; Gilead Sciences; Janssen Pharmaceuticals, Inc. and Vertex Pharmaceuticals; past partial funders include Bristol-Myers Squibb. Granting corporations do not have access to CHeCS data and do not contribute to data analysis or writing of manuscripts.

## Disclosure

Stuart C Gordon receives grant/research support from AbbVie Pharmaceuticals, Bristol-Myers Squibb, Conatus, CymaBay, Exalenz BioScience, Gilead Pharmaceuticals, Intercept Pharmaceuticals, and Merck. He is also a consultant/advisor for AbbVie Pharmaceuticals, Bristol-Myers Squibb, CVS Caremark, Gilead Pharmaceuticals, Intercept, and Merck, serves as a speaker/teacher in programs sponsored by Intercept Pharmaceuticals. The authors report no other conflicts of interest in this work.

## References

- Goldberg D, Lewis J, Halpern S, Weiner M, Lo Re V 3rd. Validation of three coding algorithms to identify patients with end-stage liver disease in an administrative database. *Pharmacoepidemiol Drug Saf.* 2012;21(7):765–769.
- Gordon SC, Pockros PJ, Terrault NA, et al. Impact of disease severity on healthcare costs in patients with chronic hepatitis C (CHC) virus infection. *Hepatology.* 2012;56(5):1651–1660.
- Holmberg SD, Lu M, Rupp LB, et al. Noninvasive serum fibrosis markers for screening and staging chronic hepatitis C virus patients in a large US cohort. *Clin Infect Dis.* 2013;57(2):240–246.
- Regev A, Berho M, Jeffers LJ, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol.* 2002;97(10):2614–2618.

5. Runyon BA; Practice Guidelines Committee, American Association for the Study of Liver Diseases (AASLD). Management of adult patients with ascites due to cirrhosis. *Hepatology*. 2004;39(3):841–856.
6. Garcia-Tsao G, Friedman S, Iredale J, Pinzani M. Now there are many (stages) where before there was one: in search of a pathophysiological classification of cirrhosis. *Hepatology*. 2010;51(4):1445–1449.
7. Spradling PR, Rupp L, Moorman AC, et al. Hepatitis B and C virus infection among 1.2 million persons with access to care: factors associated with testing and infection prevalence. *Clin Infect Dis*. 2012;55(8):1047–1055.
8. Kramer JR, Giordano TP, Soucek J, Richardson P, Hwang LY, El-Serag HB. The effect of HIV coinfection on the risk of cirrhosis and hepatocellular carcinoma in U.S. veterans with hepatitis C. *Am J Gastroenterol*. 2005;100(1):56–63.
9. Kramer JR, Davila JA, Miller ED, Richardson P, Giordano TP, El-Serag HB. The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs administrative databases. *Aliment Pharmacol Ther*. 2008;27(3):274–282.
10. Lo Re V 3rd, Lim JK, Goetz MB, et al. Validity of diagnostic codes and liver-related laboratory abnormalities to identify hepatic decompensation events in the Veterans Aging Cohort Study. *Pharmacoepidemiol Drug Saf*. 2011;20(7):689–699.
11. Nehra MS, Ma Y, Clark C, Amarasingham R, Rockey DC, Singal AG. Use of administrative claims data for identifying patients with cirrhosis. *J Clin Gastroenterol*. 2013;47(5):e50–e54.
12. SI Inc. *SAS/STAT Software*. Version 9.4. 8th ed. Cary, NC: SAS Institution Inc.; 2013.
13. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387.
14. *CART 6.0 User's Guide Salford Systems* [computer program]. 2010.
15. Breiman L, Friedman J, Stone Charles J. *Classification and Regression Trees*. 1st ed. New York: Chapman and Hall; 1984.
16. Moore CL, Lu M, Cheema F, et al. Prediction of failure in vancomycin-treated methicillin-resistant *Staphylococcus aureus* bloodstream infection: a clinically useful risk stratification tool. *Antimicrob Agents Chemother*. 2011;55(10):4581–4588.
17. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–845.
18. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York: John Wiley & Sons; 1986.

## Clinical Epidemiology

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Dovepress